Binaural models and virtual acoustics to study spatial perception

Schimmel, S ; Mueller, M ; Dillier, N

Abstract: To evaluate all aspects of binaural signal processing algorithms, an objective measure of how they affect spatial perception is required. We are developing such a measure using a data-driven approach based on virtual acoustics and models of binaural signal processing of the auditory system. We present a brief overview of the methods and models that we use to study spatial perception. We discuss the binaural cues of realistic signals in more detail, and demonstrate the importance of interaural coherence as a binaural cue. We conclude with an outlook on our future work.

BINAURAL MODELS AND VIRTUAL ACOUSTICS TO STUDY SPATIAL PERCEPTION

Steven Schimmel, Martin Mueller, Norbert Dillier

Lab. for Exp. Audiology, ORL Clinic, University Hospital, Zurich, Switzerland

Abstract

To evaluate all aspects of binaural signal processing algorithms, an objective measure of how they affect spatial perception is required. We are developing such a measure using a data-driven approach based on virtual acoustics and models of binaural signal processing of the auditory system. We present a brief overview of the methods and models that we use to study spatial perception. We discuss the binaural cues of realistic signals in more detail, and demonstrate the importance of interaural coherence as a binaural cue. We conclude with an outlook on our future work.

Background

Binaural signal processing algorithms are actively developed now that binaural hearing aids have become a reality. Binaural algorithms have two significant benefits over traditional monaural algorithms. First, they have more capabilities, because they have access to more microphone signals, and these microphones have larger spatial separation. Both aspects allow for more powerful signal processing. Second, they have full control over the binaural cues that they present to the user, and thus they have – theoretically – full control over the perception of acoustical space for the hearing aid user.

This last property of binaural algorithms creates the need for more advanced algorithm evaluation. Not only should binaural algorithms be evaluated on standard criteria such as noise reduction, speech intelligibility and signal quality, they should also be evaluated with respect to spatial perception. In our work, we recognize several aspects of spatial perception, which fall into two broad categories. The first category is the spatial perception of sound sources, consisting of source direction (i.e., azimuth and elevation), distance, and source width. This category also includes whether a sound source is perceived intact and not as multiple sound objects at different locations after processing. The second category deals with the spatial perception of the acoustical environment, and deals with aspects such as the room impression and the sense of envelopment.

It should be noted that some binaural algorithms may alter one or more aspects of spatial perception by design. For example, an algorithm may be designed to always present the target sound source directly in front, irrespective of the original location of the sound source. In such cases, we would like to be able to measure how well the algorithm achieves this objective, rather than how well it preserves the original acoustical scene.

The gold standard to evaluate algorithms is the subjective listening test. There have already been several studies measuring localization performance and externalization performance after processing with (binaural) hearing aid algorithms (see for example [1–4]). However, one drawback of these subjective tests is that they are time consuming. Usually, they are only used to evaluate or compare several finished products, and their use is limited during the algorithm development phase. Moreover, we've found from our own experience that it requires expert subjects to evaluate spatial perception, as untrained subjects typically lack the awareness and/or the vocabulary to describe spatial attributes of acoustic scenes. In our listening tests, for example, some subjects reported that they did not understand what it meant for a sound source to be externally perceived or internally perceived. They were only able to make the distinction with great difficulty. Similarly, the perception of a single source as multiple sound objects (which happens quite easily as a side-effect of frequency-dependent signal processing) and the concept of source width were confusing to the untrained subjects.

We are developing an objective measure of spatial quality as an alternative to subjective tests. The aim of the objective measure is to predict how subjective spatial perception of a signal is affected by processing with a hearing aid algorithm. Our approach to defining such an objective measure is as follows. First, we simulate binaural data for many realistic acoustical scenes. Next, we extract binaural cues from this data using an existing model of binaural signal processing of the human auditory system. In parallel, we evaluate spatial perception on the acoustical scenes using expert subjects. Finally, we train statistical models on the binaural cues to predict the expert spatial perception. In the following sections, we discuss each step in more detail.

Materials and Methods

We have developed a room acoustics simulator called ROOMSIM [5]. It is published as an open source project and freely available to anyone under the GNU General Public License [6]. It simulates only "shoebox" type rooms in order to keep the simulator simple and easy to use. To simulate a room, ROOMSIM needs a room size and surface absorption coefficients. The absorption coefficients can be easily selected from a large database of construction materials. In addition, ROOMSIM needs a sound source type and location, and a sound receiver type and location. ROOMSIM supports many types of directional sensitivities

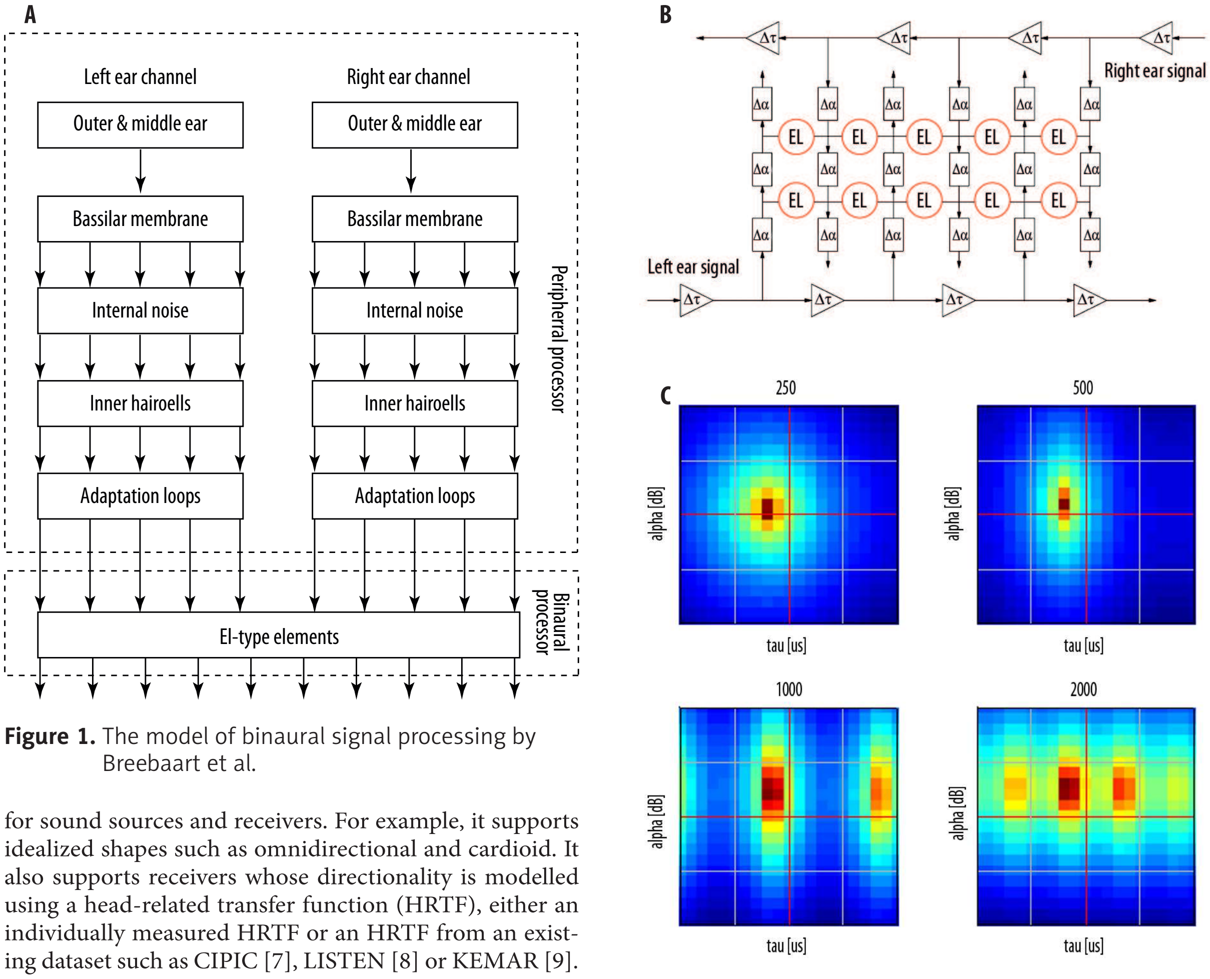

Figure 1. The model of binaural signal processing by Breebaart et al.

for sound sources and receivers. For example, it supports idealized shapes such as omnidirectional and cardioid. It also supports receivers whose directionality is modelled using a head-related transfer function (HRTF), either an individually measured HRTF or an HRTF from an existing dataset such as CIPIC [7], LISTEN [8] or KEMAR [9].

Given a monaural input signal for each sound source in the simulation, ROOMSIM simulates the room acoustics and computes a binaural output for each sound receiver in the simulation. The room acoustics are simulated using a hybrid algorithm, which is a combination of the "image source model" [10] to model the direct sound and the early reflections, and the so-called "diffuse rain" algorithm [11] to model the diffuse reflections and the reverberant tail. Once an acoustical scene is simulated, the output signal of ROOMSIM can be processed with a model of binaural signal processing to extract the binaural cues.

Initially, we used the model of binaural signal processing by Breebaart et al. [12]. This model was selected because it closely matches the physiology of the binaural auditory system, and because the authors took great care to fit the model parameters to subjective data. Moreover, it models neurons that are selective to a specific combination of interaural time and level differences (ITD and ILD). It therefore returns the largest amount of binaural information of all the models that we considered.

The Breebaart model is illustrated in Figure 1. It contains a standard front-end for peripheral processing, consisting of outer and middle ear response filters, basilar membrane response, internal noise, inner hair cell transduction, and adaptation loops (Figure 1A). The peripheral processor is connected to a binaural processor, consisting of an arrangement of EI type elements (Figure 1B). Each EI element responds to the left and right signals at a specific time ($\Delta\tau$) and level ($\Delta\alpha$) difference. The output of the EI elements for an arbitrary signal is illustrated in Figure 1C. The graph shows that for this particular signal the EI response was strongest for a slightly negative ITD and a slightly positive ILD. The full Breebaart model computes one of these graphs for each frequency band and each time instance of the input signal.

In previous work [13], we have used the Breebaart model successfully to estimate sound source direction for anechoic signals. However, it turns out that the neural selectivity of the Breebaart model is highly redundant. The ITD/ILD information provided by the EI elements is very smooth with a single peak that moves only slowly over time, even when the input signal contains multiple sound sources. In our experiments we found that the location of the peak response in ITD and ILD was the most important piece of information. We also found that the model had a weak response to reverberation, and that reflections had little effect on the model output. But perhaps most importantly, the Breebaart model does not provide a measure of interaural coherence. It turns out that coherence is essential to determine the relevance of the ITD/ILD information when the model is applied to reverberant signals, as we will see in the next section.

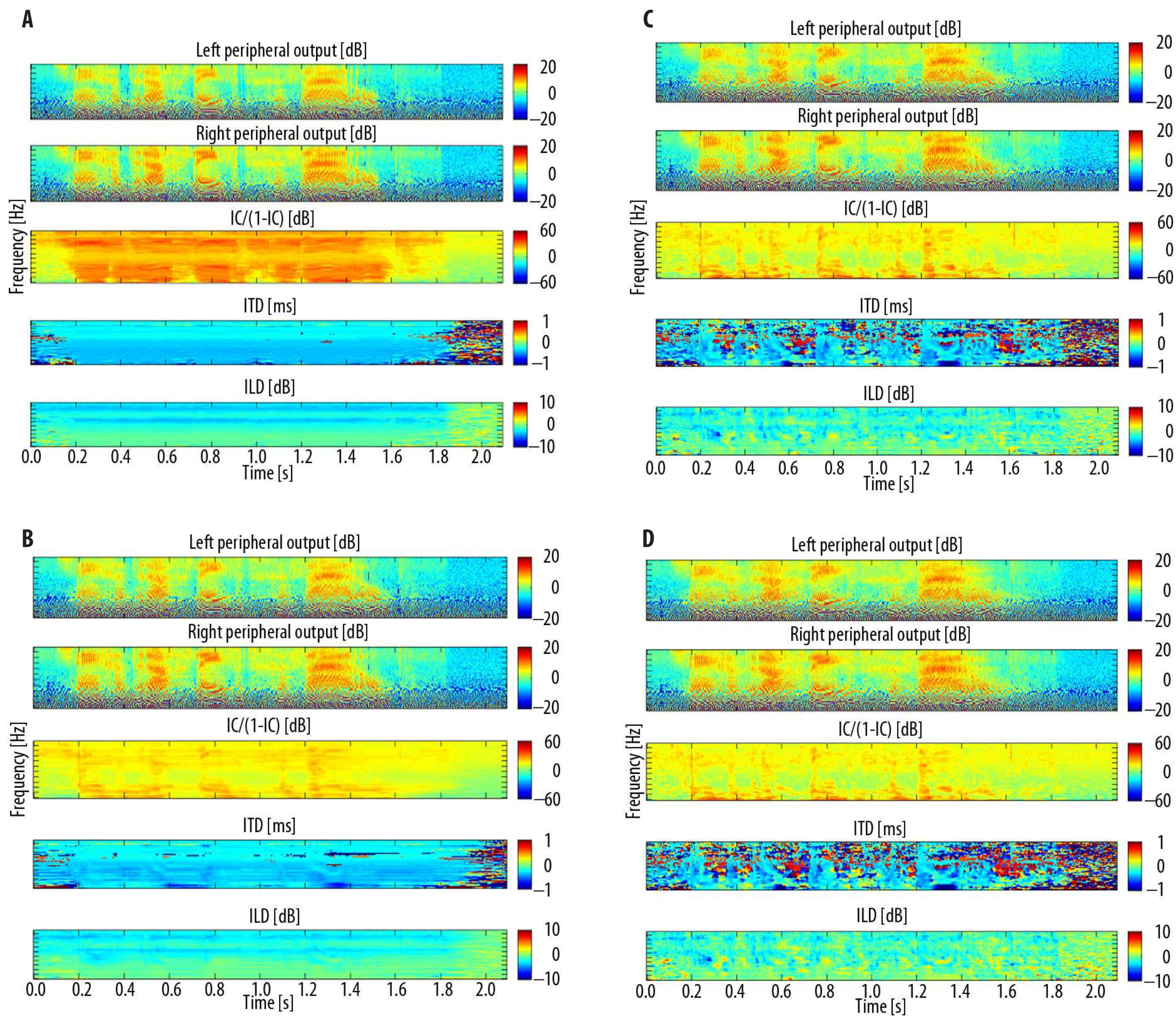

Figure 2. Output of the Faller-Merimaa model for four rooms with increasing RT60.

To extend our objective measure of spatial perception to reverberant signals, we use the model of binaural signal processing defined by Faller and Merimaa [14]. It has peripheral processing that is very similar to the Breebaart model, but a considerably different binaural processor. The binaural processor includes a measure of interaural coherence (IC) and models only the peak ITD and ILD response.

Results

The output of the Faller-Merimaa model for four simulated rooms with increasing reverberation times is illustrated in Figure 2A–D. In the simulations, the source-receiver distance was 2 meters, and the source is position 30 degrees to the right of the receiver's median plane. In each subplot, the top two panels show the left and right output of the peripheral stage. The middle panel shows the interaural coherence, shown here as IC/(1-IC) in decibels to enhance the resolution around IC=1. The bottom two panels show the interaural time and level differences. In each panel, time runs horizontally from 0 to 2 seconds, and frequency runs vertically, from 100 to 3000 Hz spaced linearly on an ERB scale.

For the anechoic condition (Figure 2A), we see that whenever there is signal power, coherence is high and ITD and ILD are nearly constant over time. When reverberation is added (Figure 2B, RT60=80 ms), the interaural coherence is now only high during the onsets of signal power, and much smaller anywhere else. The ITD and ILD values remain close to their anechoic values, but start to show more fluctuation. When reverberation time increases (Figure 2C, RT60=130 ms), these changes are even more pronounced, with significantly more noisy ITD and ILD values, Finally, for the last room (Figure 2D, RT60=170 ms) we see that these effects continue even further. From these plots it is obvious that the Faller-Merimaa model responds effectively to reverberation. Given this improved model output, the question now becomes: how do these binaural cues relate to the sound source attributes of spatial perception?

Discussion

It turns out that interaural coherence is the key to many attributes of spatial perception. As proposed by Faller and Merimaa, it is likely that the auditory system only judges sound source direction during periods of high interaural coherence [14]. From Figure 2B–D it is clear that the

ITD and ILD values at moments of high IC are more stable and closer to their anechoic values than at other times. So in order to predict the perception of sound source direction, is seems reasonable to treat ITD/ILD values during high IC with more importance than during low IC. And it seems reasonable to assume that the number, duration and strength of high IC periods are probably good predictors of the effort it will take subjects to localize a sound source.

Interaural coherence is not only important for source direction, but also for source distance. It is known that the auditory system judges source distance based on the direct-to-reverberant energy ratio, as well as on the absolute signal power of the direct sound [15]. Again, we can distinguish between direct sound and reverberance by looking at instances of high IC for direct sound and instances of low IC for reverberance. Interaural coherence can also play an important role to predict perceived auditory source width. As proposed by Mason et al. [16], the absolute strength of the interaural coherence during direct sound – that is, periods of high IC – is a measure for the perceived source width. In addition, it can be assumed that compact sources have fairly constant ITD and ILD values during direct sound, whereas the ITD and ILD values for broader sources will show more variation.

To effectively use interaural coherence as a binaural cue, we need to distinguish high IC values from low IC values. Faller and Merimaa proposed to use frequency-dependent thresholds to separate the two classes. That approach worked well in their study, but does not generalize well beyond that. Another approach, by Dietz et al. [17], is to compute the IC with frequency-dependent window lengths, so that a frequency-independent IC threshold can be used. We have adopted the latter technique, and are currently investigating statistical and data-driven approaches to classify interaural coherence in a large range of simulated acoustical scenes. In parallel, we are gathering more data on spatial perception in subjective listening tests. In the end, we will bring together the subjective perceptual data with our statistical models on the binaural cues, to predict subjective spatial perception of source direction, source distance, and source width.

References:

1. Müller M, Kegel A, Schimmel S et al: Localization of virtual sound source with bilateral hearing aids in realistic acoustic scenes. J Acoust Soc Am, 2011 (submitted)
2. Van den Bogaert T, Carette E, Wouters J: Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna. Int J of Audiology, 2011; 50: 164–76
3. Van den Bogaert T, Klasen T, Moonen M et al: Horizontal localization with bilateral hearing aids: without is better than with. J Acoust Soc Am, 2006; 119: 515–26
4. Hartmann WM, Wittenberg A: On the externalization of sound images. J Acoust Soc Am, 1996; 99: 3678–88
5. Schimmel SM, Müller MF, Dillier N: A fast and accurate 'shoebox' room acoustics simulator. Proc of ICASSP '09, 2009; 241–44
6. ROOMSIM room acoustics simulator, URL: *http://sourceforge.net/projects/roomsim/*
7. CIPIC HRTF dataset, URL: *http://interface.cipic.ucdavis.edu/sound/hrtf.html*
8. LISTEN HRTF dataset, URL: *http://recherche.ircam.fr/equipes/salles/listen/index.html*
9. KEMAR HRTF dataset, URL: *http://sound.media.mit.edu/resources/KEMAR.html*
10. Allen JB, Berkley DA: Image method for efficiently simulating small-room acoustics. J Acoust Soc Am, 1979; 66: 943–50
11. Heinz R: Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. Applied Acoustics, 1993; 38: 145–58
12. Breebaart J, van der Par S, Kohlraush A: Binaural processing model based on contralateral inhibition. I. Model structure. J Acoust Soc Am, 2001; 110: 1074–88
13. Schimmel SM, Müller MF, Dillier N: Modeling the perception of auditory space: sound source direction. Presented at the Audiological Research Cores (ARCHES) meeting, 2009
14. Faller C, Merimaa J: Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. J Acoust Soc Am, 2004; 116: 3075–89
15. Zahorik P, Brungart DS, Bronkhorst AW: Auditory Distance Perception in Humans: A Summary of Past and Present Research. Acta Acustica united with Acustica, 2005; 91: 409–20
16. Mason R, Brookes T, Rumsey F: Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. J Acoust Soc Am, 2005; 117: 1337–50
17. Dietz M, Ewert SD, Hohmann V: Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication, 2011; 53: 592–605